

We claim:

1. A computer-implemented method of processing a document, said method comprising:  
converting a document into a common format document;  
5 recognizing a concept in said common format document, wherein said concept represents  
a basic idea expressed in said common format document; and  
incorporating said concept in a conceptual model.
2. The computer-implemented method of claim 1, wherein recognizing said concept  
10 includes:  
identifying a plurality of features in said common format document, wherein said  
plurality of features represents evidence of said concept in said common format document.
3. The computer-implemented method of claim 2, wherein recognizing said concept further  
15 includes:  
calculating a concept weight for said concept using a plurality of feature weights  
associated with said plurality of features, wherein said concept weight represents a recognition  
confidence level for said concept; and  
comparing said concept weight with a predetermined threshold value.
- 20 4. The computer-implemented method of claim 1, further comprising:  
by referencing said conceptual model, generating an auto-attribute, said auto-  
attribute being a descriptive label for said common format document.
- 25 5. The computer-implemented method of claim 1, further comprising:  
by referencing said conceptual model, assigning said common format document  
to a subject category.
6. The computer-implemented method of claim 1, wherein said converting includes  
30 converting said document into a common format document that is in an XML format.

7. A computer-readable medium to direct a computer to function in a specified manner, comprising:

instructions to recognize a basic idea expressed in a document;

instructions to assign a concept identification to said basic idea; and

instructions to generate a conceptual model based upon said concept identification.

8. The computer-readable medium of claim 7, wherein said instructions to recognize said basic idea include:

instructions to determine whether a plurality of features is present in said

document, wherein said plurality of features represents evidence that said basic idea is expressed in said document.

9. The computer-readable medium of claim 8, wherein said instructions to recognize said basic idea further include:

instructions to calculate a recognition confidence level for said basic idea using a plurality of feature weights associated with said plurality of features; and

instructions to compare said recognition confidence level with a predetermined threshold value.

10. The computer-readable medium of claim 9, wherein said instructions to generate said conceptual model include:

instructions to incorporate said recognition confidence level in said conceptual model.

11. The computer-readable medium of claim 7, further comprising:

instructions to assign an auto-attribute to said document based upon said conceptual model, wherein said auto-attribute represents a descriptive label for said document.

12. The computer-readable medium of claim 7, further comprising:

instructions to place said document in a category of a categorization taxonomy based upon said conceptual model, wherein said categorization taxonomy includes a plurality of categories.

- 5     13.     The computer-readable medium of claim 12, wherein said instructions to place said document in said category include:

instructions to assign an auto-category to said document, wherein said auto-category represents a descriptive label for said category.

- 10    14.     A computer, comprising:

a processor; and

a memory connected to said processor, wherein said memory includes:

a document modeling module, said document modeling module having:

a first module configured to direct said processor to recognize a concept in a document, wherein said concept represents a basic idea expressed in said document;

and

a second module configured to direct said processor to generate a conceptual model based upon said concept.

- 20    15.     The computer of claim 14, wherein said memory further includes:

a document integration module, said document integration module having:

a third module configured to direct said processor to convert an initial format document to said document, which has a common format.

- 25    16.     The computer of claim 15, wherein said document integration module further has:

a fourth module configured to direct said processor to separate a text portion from said initial format document; and

a fifth module configured to direct said processor to incorporate said text portion in said document.

30

17.     The computer of claim 14, wherein said first module has:

a sixth module configured to direct said processor to determine whether a plurality of features is present in said document, wherein said plurality of features represents evidence of said concept in said document;

a seventh module configured to direct said processor to calculate a concept weight for said concept using a plurality of feature weights associated with said plurality of features, wherein said concept weight represents a recognition confidence level for said concept; and

an eighth module configured to direct said processor to compare said concept weight with a predetermined threshold value.

10 18. The computer of claim 14, wherein said memory further includes:

a modeling directory,

and wherein said document modeling module further has:

a ninth module configured to direct said processor to store said conceptual model in said modeling directory.

15 19. The computer of claim 14, wherein said document modeling module further has:

a tenth module configured to direct said processor to generate an auto-attribute based upon said conceptual model, wherein said auto-attribute represents a descriptive label for said document.

20 20. The computer of claim 14, wherein said document modeling module further has:

an eleventh module configured to direct said processor to categorize said document in a category of a plurality of categories based upon said conceptual model.